

Weekly Report

6th May, 2018

Done:

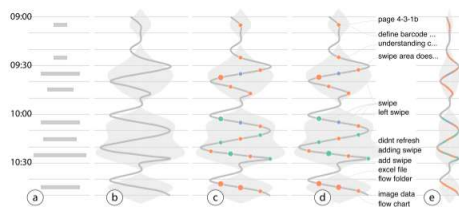
1. A post-doc conducts an English writing workshop (will lasts several weeks). I've joined the workshop and try to learn more skills.
2. Implement the polygon / multiline selection functionality and share the experience with other guys on Github. Display the time-series data with react-vis lib.
3. The descriptive data is still incomplete. Many data issues still exist and may hinder proceeding analysis. I discuss with the guy who crawls the data and tell him more about the potential issues in the sample data. Hopefully it will be ready next week, and I plan to do some basic dimension reduction or clustering.
4. Finish the online course on CNN case studies. Implement ResNet using Keras, a more high-level library for deep learning. Read the paper to learn more about its theory.
5. About the future topic of the fire data, there are two candidates after a short discussion with Prof. Ma.:
 - (1) since Max is also implementing some functionalities, there may be many interactions/visual encoding issues. Find a point and write it as a paper. It's not clear yet, I will keep in touch with Max.
 - (2) the recommendation on graph stuff I've been thinking and discussing recently. Instead of showing a graph to users, we may track the interactions of users, construct a hidden graph and do some recommendation. This is similar to Tarik&Yingcai's paper (this can also have a variation like the VIS17 Graphiti paper), and that paper lacks user study (Prof. Ma think it is not that complete result without a formal user study). We can also implement this in our system and do user studies. Perhaps we can submit this to CHI.

To Do:

1. A preliminary analysis of descriptive data once it's ready.
2. Think more concrete ideas on the two topics mentioned above.
3. As two of the functionality are partially implemented, I need to finish them and begin integrating all these subcomponents to the system.
4. Finish revising the black Visualization bible

Paper reading:

CHI18 *T-Cal: Understanding Team Conversation Data with Calendar-based Visualization* 不同时间尺度(年, 月, 天)的数据探索, 结合了 sentiment analysis 和关键词提取功能. 设计简洁, 简短几句话就可以说清问题, 也能与 design alternative 有很清楚的比较. ThreadPulse 是一种我们可以借鉴的可视设计. 比较有意思的是上面这位提到的 post-doc 正在做类似的东西, 不过是语言对话而且是实时提取.



VAST VIS17 *ACTIVIS: Visual Exploration of Industry-Scale Deep Neural Network Models* 一篇来自 facebook 的深度学习可视化的文章, 虽然标题写着 industry-scale 但其用来 demo 的数据样本看起来十分小. 其分析的基础是不同 neuron 的激活值的分析, 这个据说是普遍的操作, 但是我了解比较有限. 其一个视图几乎照搬了去年微软的分类器可视化. 系统比较简洁, 在其设定的分析任务上应该还是不错的, 但是我仍然觉得这个东西实际应用起来可能没有上次那些专门搞 DNN 的人做的有用处.

INFOVIS17 *Visualizing Nonlinear Narratives with Story Curves* 对于电影叙事手法的可视化, 核心的设计 story curve 把讲时间顺序、故事讲述顺序用 xy 两个方向展示出来. 背景色、小矩形的颜色再来编码时长、任务等信息, 十分简洁. 系统还有一些上下文信息. 通过这个工具作者发现了一些电影分析领域未指出的拍电影手法的 pattern. 这个设计虽然简洁但其任务之前确实没有做过, 这个点找的比较小而精. 在 javascript 的异步任务也可以用它来画, 但是相对而言其 pattern 可能没有那么丰富

CVPR16 *Deep Residual Learning for Image Recognition* 阅读本文主要是因为 Andrew Ng 在视频中仍没有讲清楚残差网络背后一些故事. 总的来说, 文章初心是基于网络越深理论上训练误差应该越小(不过拟合情况下), 但实际情况并非如此. 作者构造一个全等映射单独作为一层, 理论上训练出来的模型效果不会差于没有这一层, 因为我多了一层. 但是实际应用中, 这种全等映射并不能被学习出来, 于是作者提出了学习残差而不是学习本来的复杂映射的概念.

后期投稿详见 **Done** 第五点.